

child's feelings of efficacy for gendered skills and tasks. Children first acted out a story which they had been told by the researcher. They then had two free play sessions, one with the assigned doll and the other without. Lastly, participants answered questionnaires about their future occupation and family role. Though the current data was analyzed using hierarchical regression and no relationship was found between ratio scores of play types and future occupation or future family roles, data collection is still in progress. This study was funded in part by Washington and Lee University and the Virginia Academy of Science.

Statistics

SYNCHRONY OF HETEROGENEOUS SPIKING NEURON MODELS. Cheng Ly, Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, Virginia 23298. Heterogeneity is a realistic physiological attribute neglected in many mean-field models in Neuroscience. To this end, we consider a coupled stochastic neural network model where each neuron has distinct or heterogeneous intrinsic properties. A proper description of this system is large dimensional and unwieldy, requiring reduced descriptions for tractability. We present some results of reduction methods to capture various statistical quantities of interest (i.e, level of synchrony).

STATISTICAL PATTERN RECOGNITION USING GAUSSIAN COPULA. Sumen Sen & Norou Diawara, Department of Mathematics and Statistics, Old Dominion University, Norfolk, Va 23529. Statistical pattern recognition is a field of study interested in classification accuracy and feature selection. Much significant research effort has been done because such problems have a vast area of applications in automatic character recognition, medical diagnostic. Classical discrimination theory assumes normality when calculating joint and conditional distribution and estimation of parameter models. However, such normality assumption is often questionable and quite restrictive. In some situations, the pattern vector is a mixture of discrete and continuous random variables. In this talk, we use copula densities to model class conditional distribution. These types of model structures are useful for a mixed pattern vector. We use simulation to compare the performance of the copula based classifier with classical normal distribution based model.

ESTIMATING SURVIVAL FUNCTIONS THROUGH MODEL COMBINING. Lihua Chen¹ & Panayotis Giannakouros², ¹Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA 22801, ²Center for Computational Mathematics and Modeling, James Madison University, Harrisonburg, VA 22801. A model combining method is developed to estimate survival functions to account for model selection uncertainty. The weighting of models is based on the predictive performance of models. This weighting method has a connection with information theory which guarantees that up to an additive penalty term of order $1/n$, the combined estimator performs as well as the best estimator in the model list. Empirical studies

demonstrate the advantage of this method. A theoretical risk bound on the combined estimator is also obtained.

EFFECTS OF ENVIRONMENTAL CHEMICAL MIXTURES ON RISK OF NON-HODGKIN LYMPHOMA. Jenna N. Czarnota, Chris Gennings, & David C. Wheeler, Department of Biostatistics, Virginia Commonwealth University, Richmond, VA 23298. Given that humans are exposed to a multitude of environmental chemicals simultaneously, it is of particular importance to examine the relationship between chemical mixtures and disease risk. Exposure profiles may change spatially, and thus it is also necessary to consider the impact of varying exposure patterns on the effect of a chemical mixture. A weighted quantile sum (WQS) approach was used in conjunction with non-linear logistic regression to model the association of a mixture of 27 correlated environmental chemicals measured in house dust and risk of non-Hodgkin lymphoma (NHL). The data were obtained from the National Cancer Institute Surveillance Epidemiology and End Results Program NHL case-control study. Analyses were performed overall (full data set) and locally (separately at each of 4 study sites), demonstrating differences in exposure, mixture effect, and relative importance of individual chemicals. Through simulation studies, the performance of WQS regression was examined in comparison to traditional shrinkage methods in terms of sensitivity and specificity in the selection of harmful chemicals.

ADJUSTING FOR COVARIATES USING PROPENSITY SCORE METHOD WHEN AUC IS USED AS MEASURE OF TREATMENT EFFECT. Hadiza Galadima & Donna McClish, Department of Biostatistics, Virginia Commonwealth University, Richmond, VA 23298. Propensity score methods have been widely used in epidemiologic research to reduce bias in cohort studies. For continuous outcomes, the mean difference between two risk groups is a well-known measure of group effect. Another effect measure for which there has been an increased interest in the literature is the probability that a randomly selected participant in the treatment group (X) has a better result than a randomly selected participant in the comparison group (Y), i.e. $P(X > Y)$. This probability is equivalent to the area under the curve (AUC), a common measure used with receiver operating characteristic (ROC) curves to assess accuracy of medical tests. We use the method of stratification on the propensity score to estimate AUC while controlling for confounding. The adjusted AUC estimator is a weighted average of the stratum-specific AUCs. Furthermore, we provide confidence interval for the adjusted AUC. Finally, we compare the adjusted AUC with the well-known Mann-Whitney non-parametric statistic. We illustrate the methodology using a sample of adults with sickle cell disease (SCD), living in the Richmond and Tidewater areas of Virginia to estimate the effect of gender on frequency of pain due to SCD controlling for differences between groups.

PLANNING FATIGUE TESTS FOR POLYMER COMPOSITES. Caleb King¹, Yili Hong¹, Stephanie P. DeHart², & Patrick A. DeFeo², ¹Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg VA 24061 & ²Applied Statistics Group, DuPont, Wilmington, DE 19803. In this paper, we present optimal designs for cyclic fatigue testing with the goal of minimizing the asymptotic variance of a lifetime percentile at a design stress level. The designs are based on a model adapted from the

fatigue literature that is derived from assumptions regarding damage accumulation in polymer composite materials. Specifically, this model is able to incorporate aspects of the testing procedure and is more suitable for modeling of cyclic fatigue in polymer composites than the model used in the current standards. We provide a comparison between our optimal designs and the traditional designs currently in use and propose a compromise design to combine the minimum variance with a suitable number of stress levels. The effects of the design and model parameters on the asymptotic variance are studied and suggestions for good designs are presented based on the results. A simulation study is used to compare the exact and asymptotic variances of the estimated lifetime percentile at the design stress level. Finally, we conclude with a summary of the results and provide some areas for future research.

MULTI-CHANGE POINTS DETECTION IN SINGLE INDEX MODEL. Hamdy F. F. Mahmoud¹, Inyoung Kim¹ & H. Kim², ¹Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg VA 24061, ²Department of Biostatistics and Epidemiology, School of Public Health, & Institute of Public Health and Environment, Seoul National University, Seoul, Republic of Korea. Environmental health studies are of great interest in human research to evaluate the relationship between mortality and temperature. It has been shown that there is a nonlinear relationship between these two variables with a fixed number of change points for temperature. Generalized linear models with log link (GLM) or generalized additive models (GAM) have been used to describe this nonlinear relationship. The currently available methods consist of two steps: they first estimate the models and then detect change points. However, the methods for simultaneously identifying the nonlinear relationship and detecting the number of change points are quite limited. To address these limitations, a unified approach is proposed for its ability to simultaneously estimate the nonlinear relationship and detect the change points. A single index change point model (SICM) is proposed as our unified approach. A permutation-based testing procedure is also provided to detect multiple change points. SICM is compared with GLM and GAM using simulation and a real application. Simulation results suggest that our approach performs better in terms of Type I error and power and outperforms in both model fitting and change points detection. The asymptotic properties of the permutation test are showed for SICM, suggesting that the number of change points is consistent. The advantages of our approach are demonstrated using the mortality data of Seoul, Korea from 2000 to 2007. Our approach detected two change points for temperature while other approaches found only one.

A TIME-VARYING PARAMETER STATE SPACE MODEL FOR THE NIGERIAN ECONOMY. Olushina Alawale Awe¹, Ian Crandell², & Scotland Leman², ¹ Department of Mathematics, Obafemi Awolowo University, Ife, Nigeria & ² Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061. In recent years, state space models have gained tremendous popularity in statistics and econometrics especially in the modeling of time series data. In this paper, we present a time-varying parameter state space model for economic indicators using data from the Nigerian Economy. Economic indicators are mainly used for measuring economic trends. Policy makers in both advanced and developing nations make use of

economic indicators like GDP to predict the direction of aggregate economic activities. We apply the Kalman filter algorithm to obtain posterior inference on state space parameters specified from a Dynamic Linear Model (DLM), which implicitly describes the overall state of the economy. Upon estimating this new model using Nigerian economic data from 1960-2009, our model is able to detect outliers, structural breaks and historical trends in the time series considered. Our initial exploratory analysis indicates that traditional leading variables like money supply, exchange rate and capital expenditure are useful in forecasting the GDP, which is indicative of the long-run economic growth of Nigeria. Also, our method provides a full option that facilitates efficient and computationally less intensive posterior simulation through Markov chain Monte Carlo (MCMC) algorithms.

PENALIZED STEREOTYPE LOGIT MODEL FOR HIGH-DIMENSIONAL DATASETS. Qing Zhou & Kellie J. Archer, Department of Biostatistics, Virginia Commonwealth University, Richmond, VA 23298. Traditional methods for modeling ordinal data do not perform well in the presence of a high-dimensional covariate space, because traditional methods require that the number of samples is greater than the number of covariates and assumes covariate independence. The general monotone incremental forward stagewise method (GMIFS) was recently adapted to fit cumulative logit, adjacent category, and continuation ratio models, and were shown to be capable of deriving a parsimonious classifier (K Archer, J Hou). One limitation of these methods is that the proportional odds assumption may not hold for all genes in the dataset. Therefore in this paper, the GMIFS method was extended to the stereotype logit model to cope with situations when the proportional odds assumption does not hold. The method was applied to gene expression dataset to predict stage of disease.

LISA 2020: A SPARK FOR GLOBAL DEVELOPMENT. Emanuel Msemo¹ & Eric Vance², ¹Department of Biometry and Mathematics, Sokoine University of Agriculture, Morogoro, Tanzania & ²Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. For quality and significant results, researchers need to adopt methods of data collection and analysis that best suit their objectives. There is a need for researchers to collaborate with statisticians in the planning and analysis stages of research for better results. To increase the global impact of statistics, LISA (Virginia Tech's Laboratory for Interdisciplinary Statistical Analysis) is partnering with universities and individuals around the world to implement its plan to create a network of 20 new statistical collaboration laboratories in developing countries by 2020. This plan is called "LISA 2020". It involves training statisticians from developing countries through a formal procedure. When the trained statisticians return to their home countries, they will establish a statistical collaboration laboratory and interact with other researchers to help solve real-world problems in statistical practice. In this talk, I will explain my involvement in LISA 2020 at Virginia Tech, and will touch upon the process of establishing a statistical laboratory at Sokoine University of Agriculture (SUA) in Tanzania. Through this laboratory, statisticians at SUA will collaborate with researchers and provide them with support and mentorship in statistical thinking, analysis and education so they will have the skills to turn their research into tangible results for Tanzanian farmers. Our belief is that well planned research accompanied by proper analysis of data is vital to improved agricultural

productivity, which will impact the lives of Tanzanians and people throughout the world.

BIVARIATE DOUBLY-INFLATED POISSON AND RELATED REGRESSION.

Pooga Sengupta & N. Rao Chaganty, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529. Count data are common in observational scientific investigations and in many instances such as twin or crossover studies, the data consist of dependent bivariate counts. An appropriate model for such data is the bivariate Poisson distribution given in Kocherlakota and Kocherlakota (2001). However, in situations where inflated count of (0,0) occur, Lee, et al. (2009) proposed the zero-inflated bivariate Poisson distribution that accounts for the inflated count. In this research we introduced and studied a bivariate distribution that accounts for an inflated count of the (k, k) cell for some $k > 0$, in addition to the inflated count for the (0,0) cell. This bivariate doubly-inflated Poisson distribution (BDIP) is a parametric model determined by four parameters (p, \cdot). In this talk we will first discuss the distributional properties such as identifiability, moments and conditional distributions and stochastic representation of the BDIP model. Next we will discuss parameter estimation by the method of moments and maximum likelihood methods and a comparison of the methods via asymptotic relative efficiency calculations. If time permits we will also discuss the BDIP regression model that incorporates covariates into the BDIP model, and we will illustrate applicability of the BDIP regression model to analyze a subset of the Australian Health Survey data.

A ROBUST ESTIMATION OF SUFFICIENT DIMENSION REDUCTION.

Qin Wang & Hossein M. Rekabdarkolaei, Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23298. Dimension reduction and variable selection play important roles in high dimensional data analysis. *MAVE* (minimum average variance estimation) is an efficient approach proposed by Xia et al (2002) to estimate the regression mean space. However, it is not robust to outliers in the dependent variable because of the use of least-squares criterion. In this project, we propose a robust estimation based on local modal regression so that it is more applicable in practice. The efficacy of the new approach is illustrated through simulation studies and a real data analysis. Furthermore, we extend the new approach to select informative variables through shrinkage estimation.

INTRODUCING COVARIATES IN THE ATTRIBUTE-LEVEL BEST-WORST DISCRETE CHOICE MODELS.

Amanda Working & Norou Diawara, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529. Marketing research is interested in including covariates into the modeling of attribute-level best-worst discrete choice experiments. In the usual discrete choice experiments (DCE), the interest is to predict which product a consumer will choose. These types of experiments do not provide much information as to why the consumer chose the product. In an attribute-level best-worst DCE, a decision maker is provided profiles comprised of attribute-levels for each attribute describing a certain type of product and is asked to pick the pair of attribute levels that contains the best and worst. The usual models for attribute-level best-worst discrete choice experiments depend only on the attributes and attribute-levels. In the model we present, we still model the choice of

best-worst attribute-level pair, but we include covariates that are class depending. The class allocation is fitted under a mixed logit model. Such heterogeneous classification generalizes the choice impact of consumers. An example is developed and given as illustration.

LISA 2020: CREATING A NETWORK OF STATISTICAL COLLABORATION LABORATORIES. Eric Vance, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg VA 24061. To increase the global impact of statistics as a discipline useful for helping researchers answer research questions, and sponsored by a Google Research Award, LISA—The Laboratory for Interdisciplinary Statistical Analysis at Virginia Tech—is partnering with universities and individuals around the world to create a network of 20 new statistical collaboration laboratories in developing countries by 2020. LISA and its partners will educate and train statisticians from developing countries to communicate and collaborate with non-statisticians and then support these statisticians to create statistical collaboration laboratories in their home countries to help researchers, government officials, local industries, and NGOs apply statistical thinking and data science to make better decisions through data. At LISA and elsewhere, we will unlock the collaborative potential of technically sound statisticians who will in turn unlock the research potential of their collaborators and teach other statisticians to do likewise. These local research collaborations, now with the power of statistical thinking and data science open to them, will be key to improving human welfare worldwide. This talk will focus on the steps of the LISA 2020 plan, including how LISA trains statisticians to become interdisciplinary collaborators, how statistical collaboration laboratories create knowledge, and how we are building a mentoring network to assist statisticians in developing countries to enable and accelerate research by making statistics more practical in solving real world problems.

Structural Biology, Biochemistry, and Biophysics

EVOLUTION OF THE SYNUCLEIN PROTEINS: ORPHANS OR SUPERFAMILY MEMBERS. Lesley Greene, Agatha Munyanyi, John Bedford, & Zeinab Haratipour, Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, VA 23529. The synucleins are composed of three related proteins: α , β and γ . The functions and structures of these proteins in the brain remain unresolved. Computational approaches such as bioinformatics can be highly informative and advance theories about the structure and function of proteins long before lengthy experimental work is completed. Computational research can also inform and direct experimental directions. In line with this perspective we conducted bioinformatics and modeling research to facilitate understanding the structure and function of the synucleins. Using PSI-BLAST searches we identified four proteins which contain large regions of significant sequence identity and similarity with the synucleins. These proteins are an endoglucanase enzyme from *Acetorbacter pomorum*, a CRE-DUR-1 protein from a nematode, a cytochrome c from a spiral bacterium and a putative protein from the Tasmanian Devil. Three-dimensional molecular modeling indicates that these proteins share a common helix-turn-helix structure with the membrane-bound form of